TRAM to recruit the signaling adaptors MyD88 and TRIF, respectively, to specialized regions of the plasma membrane where signaling complexes are assembled. TLR2 may use a similar combination of sorting (TIRAP) and signaling (MyD88) adaptors to initiate signaling from the plasma membrane. However, for many other TLRs that signal solely through MyD88 (TLR5, 7, 8, 9, and 11) or TRIF (TLR3), there are no known sorting adaptors. A future challenge will be to understand how these TLRs sort out the signals.

REFERENCES

Chen, Z.J. (2005). Nat. Cell Biol. *7*, 758–765.

Dunne, A., and O'Neill, L.A. (2003). Sci. STKE *2003*, re3.

Fitzgerald, K.A., Rowe, D.C., Barnes, B.J., Caffrey, D.R., Visintin, A., Latz, E., Monks, B., Pitha, P.M., and Golenbock, D.T. (2003). J. Exp. Med. *198*, 1043–1055.

Flo, T.H., Ryan, L., Kilaas, L., Skjak-Braek, G., Ingalls, R.R., Sundan, A., Golenbock, D.T., and Espevik, T. (2000). Infect. Immun. *68*, 6770–6776.

Kagan, J.C., and Medzhitov, R. (2006). Cell, this issue.

McLaughlin, S., Wang, J., Gambhir, A., and Murray, D. (2002). Annu. Rev. Biophys. Biomol. Struct. *31*, 151–175.

Perera, P.Y., Mayadas, T.N., Takeuchi, O., Akira, S., Zaks-Zilberman, M., Goyert, S.M., and Vogel, S.N. (2001). J. Immunol. *166*, 574–581.

Rowe, D.C., McGettrick, A.F., Latz, E., Monks, B.G., Gay, N.J., Yamamoto, M., Akira, S., O'Neill, L.A., Fitzgerald, K.A., and Golenbock, D.T. (2006). Proc. Natl. Acad. Sci. USA *103*, 6299–6304.

Yamamoto, M., Sato, S., Hemmi, H., Uematsu, S., Hoshino, K., Kaisho, T., Takeuchi, O., Takeda, K., and Akira, S. (2003). Nat. Immunol. *4*, 1144–1150.

Picking Pyknons out of the Human Genome

Alison Meynert[1] and Ewan Birney[1,*]
[1]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom
*Contact: birney@ebi.ac.uk
DOI 10.1016/j.cell.2006.05.019

In a recent paper in PNAS, Rigoutsos et al. (2006) describe a nonrandom pattern of repeated elements, called pyknons, which are found more frequently in the 3′ untranslated regions of genes than in other regions of the human genome. Although it is unclear how pyknons might have arisen, it is possible that they may be involved in a new form of gene regulation.

For decades, computer scientists have been working on the manipulation and analysis of strings of data. Many of the techniques they have developed for analyzing large corpora of text can be easily adapted to the examination of megabases of DNA information present in the human genome. A new study reported by Rigoutsos and colleagues in a recent issue of PNAS reveals the insights gained by applying computer science to biological problems. This study investigates the relationships between short repeated elements in genic and nongenic DNA in the human genome (Rigoutsos et al., 2006) using a pattern discovery tool called TEIRESIAS (Rigoutsos and Floratos, 1998). TEIRESIAS is based on a clever computational trick that allows the identification of overrepresented or degenerate sequence patterns out of the trillions of patterns that are possible. This complex computation is made manageable by the sparsity with which such patterns occur in real genomic sequences, which enforces a restricted search space. Other methods that also make use of this sparsity are based on suffix trees (Ettwiller et al., 2005; Xie et al., 2005). TEIRESIAS's unique advantage is its two-stage process of building longer patterns from shorter ones.

Rigoutsos et al. (2006) divided the human genome into genic and nongenic regions and looked for patterns of at least 16 bases in length that occurred at least 40 times in the nongenic portion. They removed patterns that did not occur at least once in the genic regions and filtered the remaining patterns so that all genic instances were nonoverlapping. There were 127,998 of these patterns, far more than one would expect to appear by chance. They called these sequence patterns pyknons from the Greek word meaning "dense." Interestingly, these pyknons are found more often in the 3′ untranslated regions (UTRs) of genes than in other regions of the genome (7.33% nucleotide coverage compared to 3.04% in exons) (Figure 1). The spacing of pyknons in 3′ UTRs is also suggestive, with the average inter-pyknon distance falling between 18 and 22 nucleotides. Rigoutsos et al. (2006) suggest that this finding hints at the possibility of complex posttranscriptional regulation events, such as those mediated by miRNAs. Small regulatory RNA sequences are typi-

cally between 18 and 31 nucleotides (Valencia-Sanchez et al., 2006). A small number of pyknons (689, or 0.5%) were clustered together with miRNA sequences identified in Rfam, a database of known noncoding RNA genes (Griffiths-Jones et al., 2003). However, the position of pyknons is the opposite of what would normally be expected for a regulatory motif (either miRNA target sequences or DNA binding motifs). Classically, one expects a single instance of an intergenic miRNA regulating a set of transcripts through a series of genic target sites. For example, a recent study of 218 known mammalian miRNAs found 4,467 genes with targets in the 3′ UTR of their corresponding transcripts (John et al., 2004). In contrast, each pyknon is found more frequently in nongenic positions rather than inside sequences encoding transcripts. If pyknons are indeed regulatory in nature, their mode of regulation must be very different from those with which we are familiar.

A frustration for computer scientists is that although DNA sequences are easy to analyze, interpreting why a sequence pattern in a genome is nonrandom is much harder to pin down. For example, patterns that appear many times in a genome might not be functionally important. Many dispersed repeats and retrotransposed pseudogenes also generate considerable numbers of related patterns in the genome. The authors point that although nearly all pyknons (99.9%) show some overlap with repeat elements, there are at least 50,000 instances of pyknons that show no overlap with repeat elements as defined by RepeatMasker (Smit et al., 1996). However, most pyknons (90%) are found at least half of the time in repeat regions, meaning that the vast majority of pyknon instances are in classical repeats.

Dispersed repeats are not only copied around the genome but are also disrupted by subsequent translocations, inversions, and deletions, some of which are mediated by other repeat insertions. As a result, it is hard to annotate every fragment of dispersed repeats. Repeat-free regions defined by RepeatMasker suggest that repeats may not be recognizable, but it is not clear that the bases did not arise from a repeat copy. Perhaps some of the instances of pyknons in "repeat-free regions" are actually repeat fragments beyond the detection range of RepeatMasker.

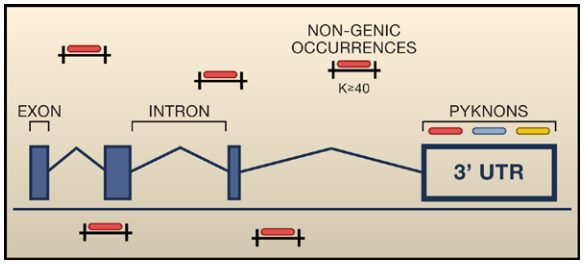

Figure 1. Pyknons in the Human Genome
Pyknons are short repeated elements in the human genome that are described by Rigoutsos et al. (2006). The pyknons in the 3′UTR of a protein-coding gene are displayed as colored bars. Some of the more than 40 occurrences of the red pyknon in nongenic regions are also depicted.

Also, the potential role of retrotransposed pseudogenes has not yet been fully addressed, although the authors state that the distances between specific pairs of pyknons is not what one would expect if the pyknons were being generated by degenerate copies of a large region. However, inserted copies of pseudogenes are usually derived from the mature mRNA, and thus the interpyknon pair distance between a real gene (with introns) and a retrotransposed pseudogene is not expected to be the same.

Additionally, the authors note that over 85% of pyknon instances in nongenic regions of the human genome are not present in the rat or mouse genome. This is circumstantial evidence that some pyknons may be repeat-related and, at the very least, not involved in the common biology of rodents and primates. Clearly a more in-depth evolutionary analysis is required.

Regardless of how pyknons are generated, the observation that they are more prevalent in 3′UTRs suggests that they may have a functional role. One observation supporting this notion is that transcripts in certain GO (Gene Ontology) categories are enriched (e.g., transcription regulation, nucleic acid metabolism) or depleted in pyknons. However, it is unclear whether this indicates a true enrichment of pyknons in these functional categories or whether it reflects the fact that longer 3′UTRs, which presumably would have a greater number of pyknons, are also known to be overrepresented in these GO categories.

There is nothing preventing biology from making use of dispersed repeats to scatter potential regulatory sites across the genome. A repeat-mediated mechanism for generating new regulatory patterns would be very exciting. One explanation for a higher rate of occurrence in 3′UTRs that is consistent with dispersed repeats is that poly(A) addition sites are often quite variable over evolutionary time and, being intron free, are more likely to incorporate dispersed repeats than 5′UTRs and coding exons. Alternatively the sequence context of 3′UTRs may preferentially attract some dispersed repeats compared to 5′UTRs.

This leaves the spacing of pyknons in 3′UTRs as the main oddity. There is no obvious reason why dispersed repeats would show a spacing of 18 to 22 nucleotides between copies. Although the evidence presented in this paper is suggestive of a new form of gene regulation, definitive tests for this hypothesis will come from future experiments that determine the effect of adding or removing pyknons from 3′UTRs.

REFERENCES

Ettwiller, L., Paten, B., Souren, M., Loosli, F., Wittbrodt, J., and Birney, E. (2005). Genome Biol. *6*, R104.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003). Nucleic Acids Res. *31*, 439–441.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). PLoS Biol. *2*, e363. 10.1371/journal.pbio.0020363.

Rigoutsos, I., and Floratos, A. (1998). Bioinformatics *14*, 55–67.

Rigoutsos, I., Huynh, T., Miranda, K., Tsirigos, A., McHardy, A., and Platt, D. (2006). Proc. Natl. Acad. Sci. USA *103*, 6605–6610.

Smit, A., Hubley, R & Green, P. (1996). RepeatMasker Open-3.0.

Valencia-Sanchez, M.A., Liu, J., Hannon, G.J., and Parker, R. (2006). Genes Dev. *20*, 515–524.

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005). Nature *434*, 338–345.

A Phosphatase Controls the Fate of Receptor-Regulated Smads

Stephen H. Schilling,[1] Michael B. Datto,[2] and Xiao-Fan Wang[1,*]
[1]Department of Pharmacology and Cancer Biology
[2]Department of Pathology
Duke University Medical Center, Durham, NC 27710, USA
*Contact: wang0011@mc.duke.edu

DOI 10.1016/j.cell.2006.05.015

In this issue of *Cell*, Lin et al. (2006) answer one of the long-standing questions in the TGFβ field by identifying a phosphatase, PPM1A, that directly dephosphorylates Smad2 and Smad3 to limit their activation.

Transforming growth factor β (TGFβ) signaling regulates numerous biological processes in a variety of cell types in organisms ranging from worms to humans (Shi and Massague, 2003). Despite the amazingly diverse set of cellular responses regulated by TGFβ, from proliferation and apoptosis to cellular differentiation and motility, the central signaling pathway downstream of TGFβ is surprisingly simple. TGFβ binds to its receptors at the cell surface, facilitating phosphorylation of the type I receptor (TβRI) by the type II receptor (TβRII). The activated TβRI then phosphorylates the receptor-activated Smads (R-Smads) Smad2 and Smad3 at two serines in their C-terminal SXS motif—a crucial step in the transduction of a TGFβ signal. Phosphorylation alters the conformation of the R-Smads, relieving their autoinhibition and releasing them from cytoplasmic retention proteins such as the Smad anchor for receptor activation (SARA) or microtubules. These conformational changes also increase the affinity of the R-Smads for the common Smad (Co-Smad), Smad4, to facilitate complex formation. The resulting Co-Smad-R-Smad complexes then translocate to the nucleus and interact with different sets of cofactors to regulate expression of specific target genes, leading to a particular biological response.

To enable the relatively simple TGFβ signaling pathway to influence such diverse biological events, many layers of tight regulation exist to control not only the level and duration of pathway activation but also what genes and responses are induced in different cell types and contexts. This regulation occurs at all levels of the TGFβ signaling pathway. The expression, bioavailability, and activation of TGFβ ligands are highly regulated, as is the expression of the type I and type II receptors. Regulation at this level dictates whether the TGFβ signaling cascade is even initiated. Once initiated, the TGFβ signal is regulated by crosstalk with other signaling pathways to either blunt or augment TGFβ-regulated transcription. Finally, the expression patterns and levels of Smads and their binding partners further fine tune transcriptional responses, allowing a diverse array of distinct cell type- and context-specific effects.

The mechanism through which these activated pathways are terminated is also a highly regulated process. Reductions in the levels of active TGFβ ligand, internalization and degradation of the TGFβ receptors, and inhibition of receptor activity through induction of the inhibitory Smads (I-Smads), Smad6 and Smad7, are all means through which the TGFβ signal is terminated upstream of the R-Smads. However, elucidating the mechanisms by which the TGFβ signal is terminated at the level of the R-Smads has proven a more difficult task. Initially, it was thought that the level of activated R-Smads in the nucleus was